

PAPER • OPEN ACCESS

# ECG derived feature combination versus single feature in predicting defibrillation success in out-of-hospital cardiac arrested patients

To cite this article: M D Ivanovi *et al* 2019 *Biomed. Phys. Eng. Express* **5** 015012

View the [article online](#) for updates and enhancements.

## Biomedical Physics &amp; Engineering Express



## PAPER

## OPEN ACCESS

RECEIVED  
17 September 2018

REVISED  
5 October 2018

ACCEPTED FOR PUBLICATION  
26 October 2018

PUBLISHED  
7 November 2018

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



## ECG derived feature combination versus single feature in predicting defibrillation success in out-of-hospital cardiac arrested patients

M D Ivanović<sup>1,2</sup> , M Ring<sup>2</sup>, F Baronio<sup>3</sup>, S Calza<sup>4</sup>, V Vukčević<sup>5</sup>, Lj Hadžievski<sup>1,6</sup>, A Maluckov<sup>1</sup> and B Eskofier<sup>2</sup>

<sup>1</sup> Vinča Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

<sup>2</sup> Machine Learning and Data Analytics Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

<sup>3</sup> CNR and Department of Information Engineering, University of Brescia, Brescia, Italy

<sup>4</sup> Unit of Biostatistics & Bioinformatics, Department of Molecular and Translational Medicine and BODAI Lab, University of Brescia, Brescia, Italy

<sup>5</sup> School of Medicine, University of Belgrade, Belgrade, Serbia

<sup>6</sup> Diasens, Belgrade, Serbia

E-mail: [marijap@vin.bg.ac.rs](mailto:marijap@vin.bg.ac.rs)

**Keywords:** classification, ventricular fibrillation (VF), defibrillation, shock outcome, feature selection, machine learning algorithms

**Abstract**

**Objective:** Algorithms to predict shock outcome based on ventricular fibrillation (VF) waveform features are potentially useful tool to optimize defibrillation strategy (immediate defibrillation versus cardiopulmonary resuscitation). Researchers have investigated numerous predictive features and classification methods using single VF feature and/or their combinations, however reported predictabilities are not consistent. The purpose of this study was to validate whether combining VF features can enhance the prediction accuracy in comparison to single feature. **Approach:** The analysis was performed in 3 stages: feature extraction, preprocessing and feature selection and classification. Twenty eight predictive features were calculated on 4s episode of the pre-shock VF signal. The preprocessing included instances normalization and oversampling. Seven machine learning algorithms were employed for selecting the best performin single feature and combination of features using wrapper method: Logistic Regression (LR), Naïve-Bayes (NB), Decision tree (C4.5), AdaBoost. M1 (AB), Support Vector Machine (SVM), Nearest Neighbour (NN) and Random Forest (RF). Evaluation of the algorithms was performed by nested 10 fold cross-validation procedure. **Main results:** A total of 251 unbalanced first shocks (195 unsuccessful and 56 successful) were oversampled to 195 instances in each class. Performance metric based on average accuracy of feature combination has shown that LR and NB exhibit no improvement, C4.5 and AB an improvement not greater than 1% and SVM, NN and RF an improvement greater than 5% in predicting defibrillation outcome in comparison to the best single feature. **Significance:** By performing wrapper method to select best performing feature combination the non-linear machine learning strategies (SVM, NN, RF) can improve defibrillation prediction performance.

**1. Introduction**

Out-of-hospital cardiac arrest (OHCA) represents one of the leading causes of death in Europe with the incidence of 86.4 per 100000 inhabitants per year. The 60% of them are treated by emergency medical service and only 9% survive to hospital discharge (Boyce *et al* 2015). The most frequent initial rhythm in OHCA is ventricular fibrillation (VF), a non-perfusing rhythm characterized by rapid and disorganized

contraction of the heart muscle cells with indiscernible ECG waveforms.

Electrical defibrillation, which consists of delivering therapeutic dose of electrical current to the fibrillating heart, is still the only effective way to revert VF and restore organized electrical activity (Sternbach *et al* 2000). Although, the early defibrillation has proved its benefit for increasing the survival rate after witnessed cardiac arrest (White *et al* 1996, Cappucci *et al* 2001), the efficiency of immediate defibrillation

for prolonged VF is debatable. The likelihood of successful defibrillation decreases rapidly with the duration of untreated VF. This is due to the increased myocardial oxygen demand during prolonged VF which results in depleted energy in the myocardium and causes a state of acidosis. Defibrillation of the heart in this state is unlikely to restore organized electrical activity (Johnson *et al* 1995). On the other hand, the myocardial condition can be improved by performing cardiopulmonary resuscitation (CPR). Clinical data indicated that survival rate in witnessed VF decreases 7% to 10% every minute of untreated VF duration, and only 3% to 4% per minute if effective CPR is provided (Link *et al* 2010).

The studies Wik *et al* (2003) and Cobb *et al* (1999) showed a crossover point of 4–5 min. After this time the initial CPR with chest compression before delivery of a defibrillation attempt improved the probability of restoring an organized electrical activity. However, in OHCA the accurate assessment of the onset time of VF is usually impossible, making it difficult to determine the priority of interventions, CPR or immediate defibrillation. Furthermore, repetitive unsuccessful defibrillation attempts with high energy are injurious to the already ischemic myocardium and can cause VF to deteriorate into asystole or pulseless electrical activity, which are difficult to resuscitate (Strohmenger 2008). Therefore, the optimal timing of defibrillation by predicting the shock outcome has become a subject of major interest.

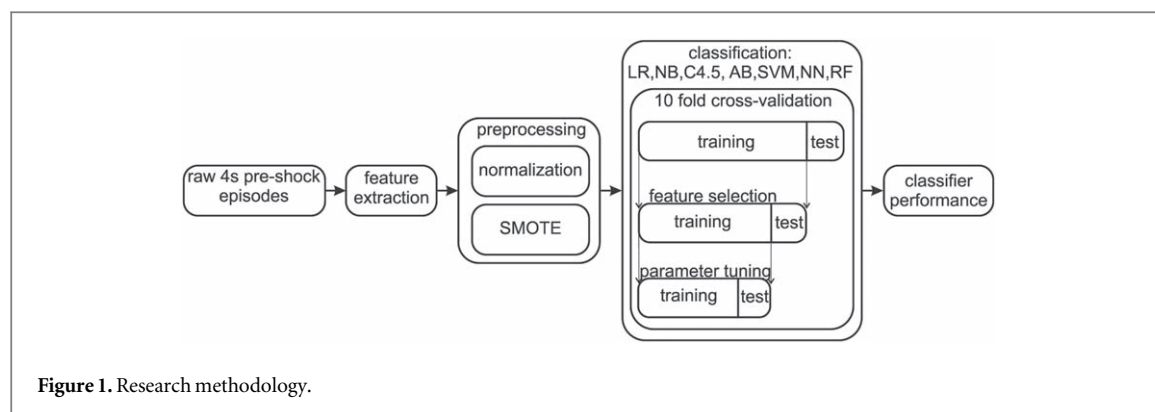
Quantitative analysis of ECG waveform, which is routinely available in the current automated external defibrillators, represents a potential tool to guide resuscitation protocols with respect to the condition of the patient. By evaluating the likelihood of a successful defibrillation outcome the optimal timing of delivering the shock can be determined. If the defibrillation attempt has a low probability of success an electrical shock should be avoided. Instead, CPR and chest compression should be utilized. Thus, the number of defibrillation attempts could be kept to a minimum. This could lead to more benefit than applying the same treatment protocol for every VF patient.

In the last two decades, different classification strategies were used to predict the defibrillation outcome of OHCA patients. They utilized single VF feature or their combinations, which were computed from the pre-shock episodes of ECG (He *et al* 2013 and references therein). To the authors' best knowledge only one publication (Shandilya *et al* 2012) addressed the issue of 'class imbalance' by using cost sensitive classification for assessing the probability of a successful defibrillation. This is an important issue to address, since many machine learning (ML) algorithms are developed with the goal of maximizing the overall accuracy. A severely imbalanced degree of accuracy can be indicated if ML algorithms are applied to imbalanced data (Chawla *et al* 2002, He and Garcia 2009).

In the study by Figuera *et al* (2016), the authors computed 30 features from each ECG segments of 4 s and 8 s and fed them to the five ML algorithms (logistic regression, bagging, random forest, boosting and support vector machine) to determine the optimal feature subset for detection of shockable rhythms. The authors reported the high performance with both sensitivity and specificity over 94% in OHCA patients. Likewise, Acharya *et al* (2018) addressed the same problem using convolutional neural networks, a deep learning algorithm which automatically discovers important features and jointly performs classification. They obtained accuracy, sensitivity and specificity over 91%. Even though, these results look promising, to determine the optimal timing of defibrillation, it is not sufficient to correctly distinguish between shockable and non-shockable rhythm. Instead, the evaluation of the probability of successful shock outcome is of major interest.

Although a few previous studies validated the usability of employing combinations of features for predicting the success of defibrillations, the reported results are not consistent. In He *et al* (2015) authors used neural networks, LR and SVM and obtained that neither combination of 16 different features nor 'best performing' or 'low correlated' feature combination can improve prediction performance in comparison to the single feature. Similarly, Neurauter *et al* (2007) asserted no improvement in predicting the defibrillation outcome using neural network and 10 features, out of which 'best predicting', 'low correlated' features or both were chosen. A study by Watson *et al* (2004) demonstrated that employing a principal component analysis (PCA) on the combination of wavelet-based features does not improve outcome prediction. They argued that PCA maximizes the variance of the classes, which does not guarantee improvement in their distinction. On the other hand, Eftestol *et al* (2000) demonstrated an improvement by combining two decorrelated PCA features. Jakova (2007) studied a set of 10 features subjected to linear discriminant analysis and showed that feature combination gave better shock success prediction. In another study, Podberger *et al* (2003) reported that using genetic programming with 3 features potentially reduces the incidence of unsuccessful defibrillations. Shandilya *et al* (2012) predicted defibrillation success by focusing on integration of multiple features (6–10) through a parametrically optimized SVM model with nested 10-fold cross-validation. Authors stated that accuracy and area under the curve were significantly improved when compared with the single feature. Additionally, Howe *et al* (2014) argued that using a SVM classification model can avoid the necessity for setting specific thresholds with adding robustness and versatility to VF waveform features.

The main objective of this study was to provide a different manner of choosing feature combination for validating whether the feature combination can



improve the prediction accuracy of defibrillation outcomes in comparison with using a single VF feature. Whereas, most of previous studies, that included combinations of features, chose combination relying only on the statistical characteristics of data without considering any particular classifier, we applied feature selection method that considers the interaction between the algorithm and the training set. The presented method deliberates a combination that provides the best synthesis instead of ranking individual features. Therefore, this method for feature selection has higher learning capacity than the ones used in previous papers. A key limitation of approaches that rank features is that they focus on the utility of individual features ignoring the influence of feature combination on the used ML algorithm. Additionally, by including the ML algorithm in feature selection process, we were able to report which ML algorithms can improve predicting performances of defibrillation success.

## 2. Methods

### 2.1. The study data

Ethical approval was granted by ethical committee of Brescia (application number NP2753). The data were part of an observational prospective study of 260 patients (>18 years old) with out-of-hospital cardiac arrest treated by the emergency medical services in Brescia, Italy, between 2006 and 2009. The patients were treated according to the 2005 European CPR guidelines (Deakin and Nolan 2005). Advanced cardiac life support management, ECG and all relevant demographic information were documented using semiautomatic Heartstart 3000 defibrillator (Laerdal Medical, Stavanger, Norway) and recorded according to the Utstein guidelines (Chamberlain *et al* 1991). The defibrillation electrodes were placed onto the patient's torso to comply with a standard lead II configuration. Patient ECGs recorded during prehospital treatment were printed in paper, scanned and converted to PDF files. These files contained more than 4 s of pre-shock and 1 min of post-shock ECGs. Afterwards they were digitized by the commercial software FindGraph for storage and offline analysis. The FindGraph software

non-uniformly sampled traces in the temporal coordinate. A 4 s episode immediately prior to the first defibrillation shock on each patient was selected for analysis and feature extraction. For preprocessing purpose, this episode was uniformly resampled to 250 Hz and band-pass filtered (0.5–48 Hz) to suppress residual baseline drift, power line interference and high frequency noise.

The post-shock rhythms were annotated independently by three experienced cardiologists who were blinded to each other's decisions. Based on the cardiologists' annotations, the shocks were categorized as successful or unsuccessful, in that the majority decision was taken. A shock was considered successful if the defibrillation returned organized electrical activity that was confirmed by ECG with the heart rate between 40–150 beats/min commencing within 1 min post-shock and persisting at least 15 s without continuing CPR. An unsuccessful shock was confirmed if VF, ventricular tachycardia, asystole, low heart rate (<40 beats/min) or pulseless electrical activity occurred after defibrillation.

### 2.2. Research methodology

The prediction analysis was performed in 3 stages (figure 1). First, the pre-shock episodes were utilized for feature extraction using Matlab 2014 software. Second, the Synthetic Minority Oversampling Technique (SMOTE) (Chawla *et al* 2002) was performed for balancing the successful and unsuccessful defibrillation class distributions using Weka software package (Hall *et al* 2009). Third, different ML algorithms were utilized for feature selection and classification via the embedded classification software toolbox (ECST) (Ring *et al* 2012).

### 2.3. Feature extraction

Twenty eight previously reported predictive features (table 1) were computed on the 4 s episode of the first pre-shock VF signals. The features can be characterized in four groups based on the signal processing techniques used for their computation.

One group of the features were obtained from time domain analysis for description of waveform amplitude, phase and slope: mean amplitude (MA), root

**Table 1.** Utilized predictive features with the references where they are defined.

Predictive features			
Time domain	Frequency domain	Time-frequency domain	Non-linear dynamic
MA (Firoozabadi <i>et al</i> 2013)	AMSA (Firoozabadi <i>et al</i> 2013)	LBEn (Endoh <i>et al</i> 2011)	MSI (Gong <i>et al</i> 2015)
RMS (He <i>et al</i> 2015)	PSA (Firoozabadi <i>et al</i> 2013)	MBEn (Endoh <i>et al</i> 2011)	SampEn (Chicote <i>et al</i> 2016)
SA (Brown and Dzwonczyk 1996)	ENRG (Firoozabadi <i>et al</i> 2013)	HBEn (Endoh <i>et al</i> 2011)	ApEn (Chicote <i>et al</i> 2016)
PPA (Firoozabadi <i>et al</i> 2013)	CF (Firoozabadi <i>et al</i> 2013)		H (Podbregar <i>et al</i> 2003)
AR (Firoozabadi <i>et al</i> 2013)	CP (Firoozabadi <i>et al</i> 2013)		ShEn (He <i>et al</i> 2015)
WA (Brown and Dzwonczyk 1996)	PF (He <i>et al</i> 2015)		DFA (Endoh <i>et al</i> 2011)
MedS (Firoozabadi <i>et al</i> 2013)	SFM (Firoozabadi <i>et al</i> 2013)		
MS (Firoozabadi <i>et al</i> 2013)	MP (Firoozabadi <i>et al</i> 2013)		
SignInt (Jekova <i>et al</i> 2004)	TE (Podbregar <i>et al</i> 2003)		
	EF (Strohmenger <i>et al</i> 1997)		

mean square (RMS), average segment amplitude (SA), average peak-to-peak amplitude (PPA), amplitude range (AR), wave amplitude (WA), median slope (MedS), mean slope (MS) and signal integral (SignInt).

Frequency domain features used for description of the frequency characteristics of VF waveforms comprised another group of predictors: amplitude spectrum area (AMSA), power spectrum analysis (PSA), energy (ENRG), centroid frequency (CF), centroid power (CP), peak frequency (PF), spectral flatness measure (SFM), max power (MP), total energy (TE) and edge frequency (EF).

The third group of features computed by the continuous wavelet transform provided concomitant spectral and temporal information: the total low-band 1–3 Hz (LBEn), mid-band 3–10 Hz (MBEn) and high-band 10–32 Hz energy (HBEn).

The fourth group of features was developed to indicate non-linear dynamical nature of VF waveforms: median stepping increment of the Poincare plot (MSI), Sample entropy (SampEn), Approximate entropy (ApEn), Shannon entropy (ShEn), Hurst exponent (H) and detrended fluctuation analysis (DFA).

## 2.4. Preprocessing

Preprocessing included instance normalization and oversampling. Since all features were numerically generated, they were normalized to the range [0, 1]. Creating new ‘synthetic’ minority class examples was performed by applying the SMOTE method (Chawla *et al* 2002). Its main idea was to create artificial examples in the minority class by interpolating between minority class examples that were close together in the feature space. In that manner, the overfitting problem which can occur by just duplicating the existing examples was avoided and the classifier created larger and less specific decision regions that generally significantly improved learning (Chawla *et al* 2002, He and Garcia 2009).

## 2.5. Feature selection and classification

Seven different ML algorithms were employed for selecting best performing feature combinations and prediction of defibrillation outcome. They included Logistic Regression (LR), Naive-Bayes (NB), Decision tree C4.5 (C4.5), AdaBoost.M1 (AB), Support Vector Machine (SVM), Nearest Neighbor (NN) and Random Forest (RF), which are briefly described in the following.

LR models posterior class probabilities using linear functions and employs a maximum likelihood estimation to learn the model parameters (Hastie *et al* 2008). Two hyperparameters were optimized during training of the classifier: number of iterations and ridge in log-likelihood<sup>7</sup>.

NB estimates a normal distribution for every class by assuming that the features are conditionally independent and normally distributed (Hastie *et al* 2008).

C4.5 constructs a decision tree to hierarchically classify instances. The tree is learned using the information gain measure (Quinlan 1993).

AB combines several weak classifiers to build a final strong one (Freund and Schapire 1996). We used decision stumps, which evaluate a single feature value based on a fixed threshold, as weak classifiers. The individual decision stumps were then weighted and combined into the final classifier. Two hyperparameters were optimized during training of the classifier: weight pruning threshold and number of stumps<sup>8</sup>.

SVMs construct a separating hyperplane which maximizes the distance between two classes (Cortes and Vapnik 1995). We used the radial basis function kernel for mapping feature vectors into a higher dimensional feature space before learning the separating hyperplane. Two hyperparameters were optimized during training of the classifier: slack penalty C, which

<sup>7</sup> The number of iterations was searched using logarithmical scale from  $10^1$  till  $10^4$ , whereas the ridge was searched in the space  $[2^{-5}, 2^5]$ .

<sup>8</sup> The weight pruning threshold was searched in range from 50 to 100 with the step of 10, whereas the number of stumps was searched over the space  $[2^1, 2^6]$ .



**Table 2.** Performance metrics of feature combination and best single feature.

Learning algorithm		Feature	A [%]	AUC [%]	Se [%]	Sp [%]	P [%]	NPV [%]
LR	combination	7	70.0	70.0	71.3	68.7	69.5	70.5
	best single	MSI	72.6	72.6	73.3	71.8	72.2	72.9
NB	combination	7	71.5	76.3	71.8	71.3	71.4	71.6
	best single	ApEn	72.1	77.3	69.7	74.4	73.1	71.1
C4.5	combination	11	73.6	76.0	82.6	64.6	70.0	78.8
	best single	AMSA	72.6	73.3	90.3	54.9	66.7	84.9
AB	combination	5	73.3	73.3	81.0	65.6	70.2	77.6
	best single	WA	72.8	72.8	85.6	60.0	68.2	80.7
SVM	combination	12	81.5	81.5	83.6	79.5	80.3	82.9
	best single	MSI	71.3	71.3	73.8	68.7	70.2	72.4
NN	combination	17	81.8	81.8	87.7	75.9	78.4	86.0
	best single	PF	76.2	83.3	57.9	94.4	91.1	69.2
RF	combination	15	82.8	82.8	90.3	75.4	78.6	88.6
	best single	PF	76.4	76.4	59.5	93.3	89.9	69.7

controls the trade-off between overfitting and generalization and the width  $\gamma$  of the kernel function<sup>9</sup>.

NN assigns the class label that is most prominent among a feature vector's  $k$  nearest neighbors (Hastie *et al* 2008). We set  $k$  to 1 and used the Euclidian distance to determine the nearest neighbors.

RFs employ the concept of bagging to synthesize multiple individual decision trees into one final classifier (Breiman 2001). Two hyperparameters were optimized during training of the classifier: the number of individual trees and maximum depth of the tree<sup>10</sup>.

For execution of the above mentioned algorithms we used the ECST, a free Java software toolbox (<http://tinyurl.com/ecstproject>). This software uses libraries from existing software packages like WEKA and LibSVM (Chang and Lin 2011).

For evaluating the performance of each algorithm, a 10 fold cross-validation was performed (Hastie *et al* 2008). Each algorithm was trained after performing further, nested cross-validation to select feature combinations. If necessary, a second nested cross-validation was performed to determine the algorithm's hyperparameters (figure 1). Feature selection was performed utilizing the wrapper method, which uses the classifier's accuracy for every candidate feature set as a performance measure. This was iteratively performed through the feature space until the specific requirements were reached. We used best-first search with a forward direction to search through the space of features by greedy hill-climbing augmented with a backtracking of 10 features. Grid search algorithm was performed to determine the values of the hyperparameter pairs stated above. This algorithm exhaustively considers all hyperparameter combinations along with a 10-fold cross-validation. It utilized the accuracy as an evaluation metric for finding the 'best' point in

the grid. After the optimal hyperparameters and feature subset were chosen, the whole training set of the outer cross-validation procedure was used to construct the classifier. The classifier was then evaluated on the holdout independent test set that was not used during the search. In this manner we are mimicking the application of the classifier to a completely independent test set. The reported performances were then calculated as the averaged values over all holdout sets of the outer 10 fold cross-validation.

Regarding the decision for the final feature subset, we employed the commonly used approach of training the entire machine learning system again (including feature selection) on the entire data set. This ensured that as much information as possible was incorporated into the determination of the final feature subset and should lead to the best possible feature selection giving the present data set.

Apart from average accuracy ( $A = TP + TN / (TP + TN + FP + FN)$ ), sensitivity ( $Se = TP / (TP + FN)$ ), specificity ( $Sp = TN / (FP + TN)$ ), precision ( $P = TP / (TP + FP)$ ), negative predictive value ( $NPV = TN / (FN + TN)$ ) and the area under the curve (AUC) of a receiver operating characteristic (ROC) plot were calculated. TP, TN, FP and FN represent the true positives, true negatives, false positives and false negatives, respectively, with positives corresponding to successful defibrillation.

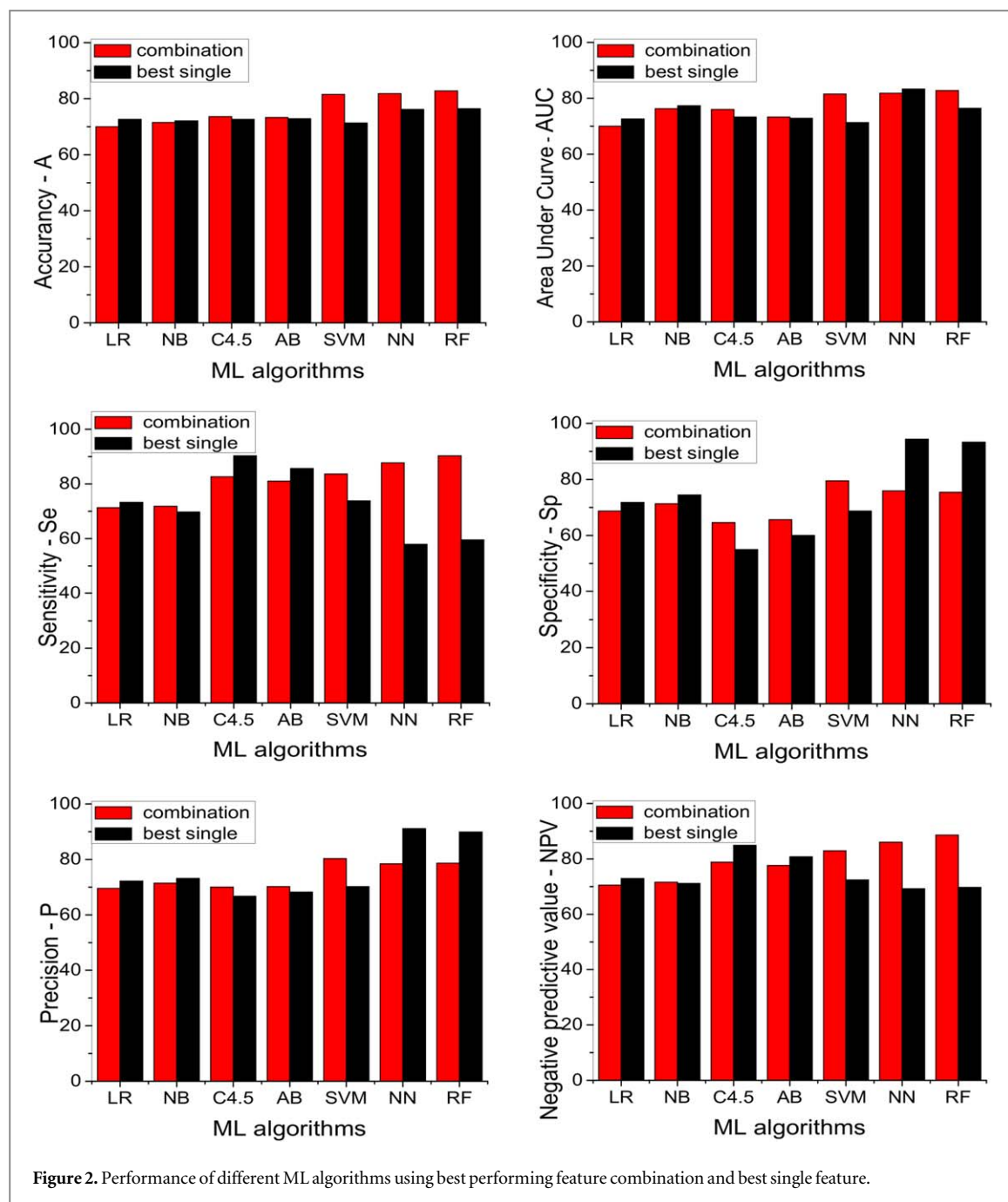
### 3. Results

By labelling the shocks, 9 of the original 260 first shocks were considered indeterminable and discarded from the analysis. Thus, 251 valid first shocks were considered, out of which 195 were unsuccessful and only 56 successful. After performing SMOTE method the classes were balanced with 195 instances in each class.

Table 2 and figure 2 show the algorithm performances for the best performing feature combination and the best single feature. The best single feature was

<sup>9</sup> Both slack penalty and width were searched over the space  $[2^{-3}, 2^6]$ .

<sup>10</sup> The number of individual trees was searched in range from 50 to 100 with the step of 10, whereas for the maximum depth we used two values 30 and 100.



determined as the feature that leads to the highest average (cross-validated) accuracy A. If more than one features had the same average accuracy, the best feature was selected as the one with the highest AUC.

The improvement of shock outcome predictive power using feature combination instead best single feature is illustrated below.

LR classifier showed no improvement in any performance measures. Using NB, Se and NPV were increased from 69.7% and 71.1% to 71.8% and 71.6%, respectively. C4.5 and AB demonstrated an improved performance for A, AUC, Sp, and P. SVM classifier showed improvement in all performance measures. Performance measures A, Se and NPV were increased when using NN and RF. AUC was also improved from 76.4% to 82.8% when RF classifier is applied (figure 2).

Typically, there is a trade-off between Se and Sp, as well as, between Se (recall) and P. Comparing different algorithms by using metrics with opposite trends could be unambiguous. Since we used accuracy to determine the best single feature and to evaluate the performance of feature combination, we have compared the proposed algorithms using A. We are aware that in case of imbalanced dataset the overall accuracy does not provide adequate information, but since we balanced our dataset by performing SMOTE the accuracy is reasonable (fair) choice for performance metric. The results showed that the linear learning algorithms (LR and NB) performed worse in predicting defibrillation outcome by using feature combination in comparison with the best feature. The reduction in A was less than 3%. By using non-linear

learning algorithms (C4.5, AB, SVM, NN and RF) an improvement in predictive power was seen. Decision tree C4.5 and AB showed increase in A not greater than 1% whereas SVM, NN and RF demonstrated increase greater than 5%. On our dataset the highest A was obtained by using the RF algorithm and feature combination. Nevertheless this algorithm had similar predictive power as SVM and NN when the best performing feature combination was used.

The number of features in the best performing subset for each ML algorithm is given in table 2. Since the twice-nested version of cross-validation would not choose the same subset in each iteration, the reported feature subsets were obtained when the whole dataset was used for feature selection. In that case, the best performing feature combination in RF algorithm contained 15 out of 28 features: RMS, SA, WA, AR, PPA, PF, MP, EF, TE, H, ShEn, DFA, LBEn, MBEn and HBEn. This feature combination achieved an accuracy of 82.8% and, therefore, increased accuracy by 6.4% compared to the best result using a single feature only.

## 4. Discussion

The major aspect of this study was to assess ML algorithms for which combinations of VF features can enhance the predictive power of the defibrillation success. Employing two levels of nesting for selection of optimal hyperparameters and features do not guarantee that the same subset of features will be selected in each iteration of external cross-validation loop. If the whole dataset is used for selecting a unique feature combination, the performances of the algorithms are optimistic, since the test sets of the external cross-validation are 'already seen' by the algorithm. Even though the unique feature subset was not utilized, the nested approach represents the only correct way for obtaining the results that are not biased by information in the test set. Therefore, the obtained results can demonstrate how the used classifiers will perform in the future when new data are presented.

In the contrast to the previous studies we have used the wrapper feature selection method, which utilizes the classifier to evaluate selected feature subsets. The literature shows that wrapper methods tend to achieve superior accuracy compared to filter methods (Gutlein *et al* 2009). In filter methods features are ranked by their relevance (the higher relevance, the greater predictive ability). Whereas feature subsets obtained by the wrapper method do not necessarily contain the individually best-ranked features, but the combination that provides the best synthesis (Kohavi and John 1997). We presume that previously reported absence of improvement in applying non-linear ML algorithms and feature combinations could be due to the selection of features by ranking them and ignoring the influence of feature combination. For instance, in He *et al* (2015) it was reported that using a SVM

accuracy does not exceed 75% regardless of whether all 16 features, combination of 'best performing' or 'low correlated' features was utilized, whereas we obtained 81.5%. The above given comparison should be taken with reserve, since the direct comparison of our results with the previously published studies is not possible, due to the usage of different datasets, definitions of defibrillation success, types of utilized defibrillators (monophasic/biphasic) and duration of pre-shock VF rhythms. Additionally, our study differs from the previous ones in that they were all focused on reporting the utility of features without paying attention to the choice of ML algorithm. On the other hand, we are focused on reporting the utility of the ML algorithms for improving the predicting performances of the defibrillation outcome.

Whereas wrapper methods obtain higher performances filter methods are faster. This drawback of wrapper methods can be overcome by implementing the idea of an offline-online system as reported in Faust *et al* (2018). In the offline stage of the system optimal hyperparameters and features can be selected based on the labeled data. Therefore, the feature combination that could give higher performances would be selected without speed concern. To validate this idea a larger amount of data (with an independent test set) is needed.

In our original dataset the number of successful defibrillations was approximately five times lower in comparison with the number of unsuccessful ones. This imposed the necessity for applying a method to deal with imbalanced learning problem. Several methods that address this issue have been developed, with sampling techniques and cost sensitive learning being most widely used. In our approach we adopted the SMOTE oversampling method, since reducing a large number of instances from the majority class might lead to omission of some important information pertaining to the majority class (Chawla *et al* 2002, He and Garcia 2009). An alternative way would be to use cost sensitive classification, which utilizes misclassification error costs. But since in our study the real costs were unknown setting a cost value that is not optimized could bias classification toward either class.

We list a few limitations of the study: SMOTE method could lead to over-generalization and artificially induced variance; VF patients were treated with a monophasic defibrillator; this study was a retrospective study, so the feature extraction and classification were performed offline; even though we performed imbalance correction and 10 fold cross-validation an independent test set was omitted; we used the most common 28 features even though there are other not so commonly used features designed for the same purpose.

Recent studies have shown that deep learning performs better than the conventional machine learning approaches. The main advantage of the deep learning approach is the ability to automatically discover useful



features from raw signals, which typically leads to higher performance. To the best of our knowledge, the deep learning strategy has not been used yet for predicting defibrillation success.

## 5. Conclusion

This paper addressed defibrillation outcome prediction using VF waveform features processed with machine learning algorithms. It was found that multiple features, which were selected with respect to the subsequent classifier (wrapper method), can improve defibrillation outcome predictions. This was particularly true if machine learning algorithms that are able to learn complex and nonlinear decision boundaries were employed, e.g., Random Forests, Nearest Neighbor or Support Vector Machines. Moreover, the typical class imbalances (successful versus unsuccessful defibrillation outcomes) were considered so that the present results should realistically reflect the algorithms' performance in real-world applications. Regarding future work, the present findings motivate in-depth exploration of deep learning methods for defibrillation outcome prediction as deep learning couples features and classification even more tightly, which could again increase performance.

## Acknowledgments

This work was financially supported in part by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 691051 and the Ministry of Education, Science and Technological Development of Serbia (III45010). B E was supported by the German Research Foundation within the framework of the Heisenberg professorship programme (grant No. ES 434/8-1). M R was supported by the Federal Ministry of Education and Research of Germany (grant No 01IS17070).

## ORCID iDs

M D Ivanović  <https://orcid.org/0000-0002-7218-6575>

## References

- Acharya U R, Fujita H, Oh S L, Raghavendra U, Tan J H, Adam M, Gertych A and Hagiwara Y 2018 Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network *Future Gener. Comput. Syst.* **79** 952–9
- Boyce L W, Vliet Vlieland T P M, Bosch J, Wolterbeek R, Volker G, van Exel H J, Heringhaus C, Schali J M J and Goossens P H 2015 High survival rate of 43% in out-of-hospital cardiac arrest patients in an optimised chain of survival *Neth Heart J* **23** 20–5
- Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
- Brown C G and Dzwonczyk R 1996 Signal analysis of the human electrocardiogram during ventricular fibrillation: frequency and amplitude parameters as predictors of successful countershock *Ann Emerg Med* **27** 184–8
- Capucci A, Aschieri D, Piepoli M F, Bardy G H, Ionomu E and Arvedi M 2002 Tripling survival from sudden cardiac arrest via early defibrillation without traditional education in cardiopulmonary resuscitation *Circulation* **106** 1065–70
- Chamberlain *et al* 1991 Recommended guidelines for uniform reporting of data from out-of-hospital cardiac arrest: the 'Utstein, style' prepared by a task force of representatives from the European resuscitation council American Heart Association, Heart and Stroke Foundation of Canada, Australian Resuscitation Council *Resuscitation* **22** 1–26
- Chang C-C and Lin C-J 2011 LIBSVM: a library for support vector machines *ACM Trans Intell Syst Technol* **2** 1–27
- Chawla N V, Bowyer K W, Hall L O and Kegelmeyer W P 2002 SMOTE: synthetic minority over-sampling technique *J. Artif. Intell. Res.* **16** 321–57
- Chicote B, Irusta U, Alcaraz R, Rieta J J, Aramendi E, Isasi I, Alonso D and Ibaguren K 2016 Application of entropy-based features to predict defibrillation outcome in cardiac arrest *Entropy* **18** 313
- Cobb L A, Fahrenbruch C E, Walsh T R, Copass M K, Olsufka M, Breskin M and Hallstrom A P 1999 Influence of cardiopulmonary resuscitation prior to defibrillation in patients with out-of-hospital ventricular fibrillation *J. Am. Med. Assoc.* **281** 1182–8
- Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97
- Deakin C D and Nolan J P 2005 European resuscitation council guidelines for resuscitation 2005. Electrical therapies: automated external defibrillators, defibrillation, cardioversion and pacing *Resuscitation* **67S1** S25–37
- Eftestol T, Sunde K, Aase S O, Husoy J H and Steen P A 2000 Predicting outcome of defibrillation by spectral characterization and nonparametric classification of ventricular fibrillation in patients with out-of-hospital cardiac arrest *Circulation* **102** 1523–9
- Endoh H, Hida S, Oohasi S, Hayashi Y, Kinoshita H and Honda T 2011 Prompt prediction of successful defibrillation from 1-s ventricular fibrillation waveform in patients with out-of-hospital sudden cardiac arrest *J Anesth* **25** 34–41
- Faust O, Hagiwara Y, Hong T J, Lih O S and Acharya U R 2018 Deep learning for healthcare applications based on physiological signals: a review *Comput Meth Prog Biomed* **161** 1–13
- Figuera C, Irusta U, Morgado E, Aramendi E, Ayala U, Wik L, Kramer-Johansen J, Eftestol T and Alonso-Atienza F 2016 Machine learning techniques for the detection of shockable rhythms in automated external defibrillators *Plos One* **11** e0159654
- Firoozabadi R, Nakagawa M, Helfenbein E D and Babaeizadeh S 2013 Predicting defibrillation success in sudden cardiac arrest patients *J Electrocardiol* **46** 473–9
- Freund Y and Schapire R E 1996 Experiments with a new boosting algorithm *Proc 13th Int Conf on Machine Learning (Bari, Italy)* pp 148–561-55860-419-7
- Gong Y, Lu Y, Zhang L, Zhang H and Li Y 2015 Predict defibrillation outcome using stepping increment of poincare plot for out-of-hospital ventricular fibrillation cardiac arrest *BioMed Res Int* **2015** 493472
- Gutlein M, Frank E, Hall M and Karwath A 2009 Large-scale attribute selection using wrappers *IEEE Symp. on computational Intelligence and Data Mining (CIDM 2009)* pp 332–9 In
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P and Witten I H 2009 The WEKA data mining software: an update *SIGKDD Explor News* **11** 10–8
- Hastie T, Tibshirani R and Friedman J 2008 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd ed (New York: Springer)
- He H and Garcia E A 2009 Learning from imbalanced data *IEEE Trans. Knowl. Data Eng.* **21** 1263–84

- He M, Chen B, Gong Y, Wang K and Li Y 2013 Prediction of defibrillation outcome by ventricular fibrillation waveform analysis: a clinical review *J Clin Experiment Cardiol* **S10** 009
- He M, Gong Y, Li Y, Mauri T, Fumagalli F, Bozzola M, Cesana G, Latini R, Pesenti A and Ristagno G 2015 Combining multiple ECG features does not improve prediction of defibrillation outcome compared to single features in a large population of out-of-hospital cardiac arrests *Crit Care* **19** 425
- Howe A, Escalona O J, Di Maio R, Massot B, Cromie N A, Darragh K M, Adgey J and McEneaney D J 2014 A support vector machine for predicting defibrillation outcome from waveform metrics *Resuscitation* **85** 343–9
- Jekova I, Mougeolle F and Valance A 2004 Defibrillation shock success estimation by a set of six parameters derived from the electrocardiogram *Physiol. Meas.* **25** 1179–88
- Jekova I 2007 Shock advisory tool: detection of life-threatening cardiac arrhythmias and shock success prediction by means of a common parameter set *Biomed. Signal Process. Control* **2** 25–33
- Johnson B A, Weil M H, Tang W, Noc M, McKee D and McCandless D 1995 Mechanisms of myocardial hypercarbic acidosis during cardiac arrest *J. Appl. Physiol.* **78** 1579–84
- Kohavi R and John G H 1997 Wrappers for feature subset selection *Artif. Intell.* **97** 273–323
- Link M S, Atkins D L, Passman R S, Halperin H R, Samson R A, White R D, Cudnik M T, Berg M D, Kudenchuk P J and Kerber R E 2010 Part 6: electrical therapies automated external defibrillators, defibrillation, cardioversion, and pacin 2010 American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care *Circulation* **122** S706–19
- Neurauter A et al 2007 Prediction of countershock success using single features from multiple ventricular fibrillation frequency bands and feature combinations using neural networks *Resuscitation* **73** 253–63
- Podbregar M, Kovacic M, Podbregar-Mars A and Brezocnik M 2003 Predicting defibrillation success by 'genetic' programming in patients with out-of-hospital cardiac arrest *Resuscitation* **57** 153–9
- Quinlan J 1993 *C4.5: Programs for Machine Learning* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.)
- Ring M, Jensen U, Kugler P and Eskofier B 2012 Software-based performance and complexity analysis for the design of embedded classification systems *Proc 21st Int Conf on Pattern Recognition ICPR 2012 (Tsukuba, Japan)* pp 2266–9
- Shandilya S, Ward K, Kurz M and Najarian K 2012 Non-linear dynamical signal characterization for prediction of defibrillation success through machine learning *BMC Med Inform Decis Mark* **12** 116
- Sternbach G L, Varon J and Fromm R E 2000 The resuscitation greats. Claude Beck and ventricular defibrillation *Resuscitation* **44** 3–5
- Strohmenger H U, Lindner K H and Brown C G 1997 Analysis of the ventricular fibrillation ECG signal amplitude and frequency parameters as predictors of countershock success in humans *Chest* **111** 584–9
- Strohmenger H U 2008 Predicting defibrillation success *Curr Opin Crit Care* **14** 311–6
- Watson J N, Uchaipichat N, Addison P S, Clegg G R, Robertson C E, Eftestol T and Steen P A 2004 Improved prediction of defibrillation success for out-of-hospital VF cardiac arrest using wavelet transform methods *Resuscitation* **63** 269–75
- White R, Asplin B, Bugliosi T and Hankins D 1996 High discharge survival rate after out-of-hospital ventricular fibrillation with rapid defibrillation by police and paramedics *Ann Emerg Med* **28** 480–5
- Wik L, Hansen T B, Fylling F, Steen T, Vaagenes P, Auestad B H and Steen P A 2003 Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation: a randomized trial *J. Am. Med. Assoc.* **289** 1389–95